BEYOND COLOR REMAPPING: PHOTOMETRIC–DETAIL NETWORK FOR LOW-LIGHT IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Lookup table (LUT)-based enhancement provides an explicit dark-to-bright mapping through indexing and interpolation. However, existing paradigms remain constrained by two key limitations. RGB-only indices cannot distinguish pixels with similar chromaticity but different illumination and contrast conditions, whereas channel-independent spatial lookup neglects cross-channel structural coherence. To address these limitations, we propose the Photometric-Detail Network (PD-Net), a two-stage framework that explicitly models photometric degradation and spatial detail separately. First, we develop a Dual-Condition Photometric LUT (DCP-LUT) that integrates RGB values, an illumination map, and a dual-scale contrast map into a 5D degradation-aware index. This index enables joint correction of brightness, color, and contrast. Second, we introduce a Channel-Pair Spatial Detail LUT (CSD-LUT) that improves texture consistency through channel-pair projection, rotation-cascaded dilated L-shaped lookup, and cyclic Gauss-Seidel propagation. Furthermore, we factorize each 5D LUT into a sum of rank-one tensors using canonical polyadic (CP) decomposition, reducing storage complexity from $O(N^5)$ to $O(KN)$ without explicitly materializing dense grids. Evaluations on nine benchmarks under full-reference and no-reference protocols show that PD-Net achieves state-of-the-art performance while retaining a compact memory footprint and explicit lookup-based inference.

1 INTRODUCTION

Low-light image enhancement aims to restore visibility, color fidelity, and structural detail in underexposed images, supporting applications including mobile photography, nighttime surveillance, and autonomous driving (Wang et al., 2025). However, limited photon counts and high sensor gain introduce coupled degradations, including low contrast, color distortion, noise, and detail loss. Deep learning has advanced this task through Retinex models (Fan et al., 2025), illumination-aware Transformers (Liu et al., 2023), Mamba architectures (Meng et al., 2026), and diffusion models (Huang et al., 2025). Yet these models require dense neural computation and lack explicit, controllable pixel-level mappings, as illustrated in Figure 1. This gap motivates a core research question: **Can we devise a model that can clearly model the pixel relationship between dark and bright domains, achieving efficient image enhancement** ?

Recently, lookup tables (LUTs) map degraded pixels to enhanced values through indexing and interpolation, providing an efficient and interpretable alternative. Since image-adaptive 3D LUTs were introduced for image enhancement (Zeng et al., 2020), studies have extended LUTs from global color transformations to learnable, content-adaptive operators. One line of research improves representation and storage efficiency through implicit modeling, attention-based fusion, and low-rank factorization (Conde et al., 2024; Fu et al., 2024; Kim et al., 2025). For example, SVDLUT (Kim et al., 2025) reduces spatial LUT redundancy through low-dimensional factorization. Another line incorporates degradation cues and structural priors (Li et al., 2023; He et al., 2026; Lin et al., 2025). For example, DPLUT (Lin et al., 2025) combines a light-adjustment LUT with a diffusion-prior-guided denoising LUT for unsupervised enhancement. Collectively, these studies show that informative indices, priors, and spatial modeling can extend LUTs beyond global color remapping.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

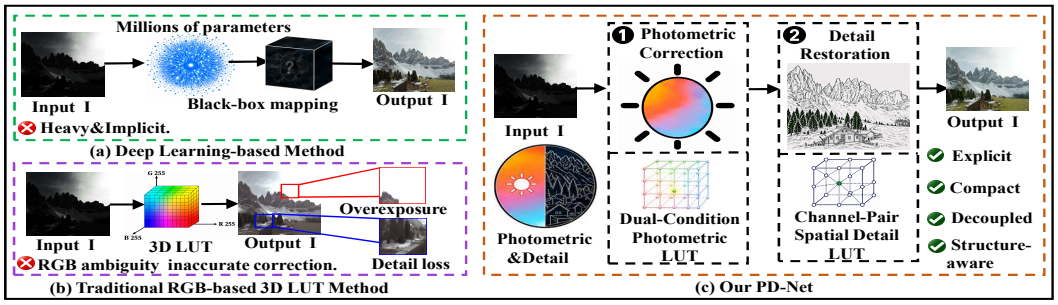


Figure 1: Beyond black-box mapping and RGB-only lookup: explicit photometric-detail modeling.

Despite these advances, existing LUT methods still face two key limitations. First, their mapping accuracy depends on whether the indices can sufficiently characterize the degradation state. With RGB-only indexing, pixels with similar colors but different illumination and contrast cannot receive targeted corrections. Additional dimensions are effective only when they encode information beyond pointwise RGB values, otherwise, a nominally high-dimensional LUT remains confined to a low-dimensional color manifold (Cai et al., 2026). Second, texture restoration requires both spatial neighborhoods and cross-channel correlations. Genuine structures usually exhibit stronger and more stable cross-channel consistency than noise, providing a cue to separate structure from noise (Yang et al., 2025). However, directly using multichannel neighborhood samples as indices increases LUT dimensionality and storage costs, whereas independent channel-wise lookup controls model size but discards cross-channel structural information. To address these limitations, we propose PD-Net, a two-stage CP-factorized LUT framework for low-light image enhancement that performs degradation-aware photometric mapping followed by channel-correlated spatial refinement.

The contributions are summarized as follows:

- We propose PD-Net, a two-stage LUT framework that decomposes low-light enhancement into degradation-aware photometric correction and channel-correlated spatial detail restoration. This lookup pipeline extends LUT methods beyond pointwise color remapping.
- We propose DCP-LUT, which builds degradation-aware 5D coordinates by integrating RGB inputs, context-guided illumination, and adaptively fused dual-scale contrast maps. This enables visually identical RGB samples under different degradations to receive distinct corrections of brightness, color, and contrast.
- We design CSD-LUT for spatial detail restoration. It projects complementary channel pairs into target-channel responses, preserving cross-channel cues without increasing spatial LUT dimensionality. Rotation-cascaded dilated L-shaped lookup and cyclic Gauss-Seidel propagation aggregate context and refine structures within each residual stage.
- We introduce a unified CP-factorized representation for photometric and spatial 5D LUTs, reducing storage complexity from $O(N^5)$ to $O(KN)$ without dense-grid materialization or interpolation changes.

2 Methodology

2.1 Overview

Given a normalized low-light RGB image I with height H, width W, and three color channels, PD-Net produces an enhanced image $\hat{I}$ through a two-stage lookup pipeline. As shown in Figure 2, the pipeline comprises the Dual-Condition Photometric LUT (DCP-LUT) for photometric correction and the Channel-Pair Spatial Detail LUT (CSD-LUT) for spatial detail restoration. DCP-LUT combines the RGB channels, illumination map D, and dual-scale contrast map C into a five-dimensional degradation-aware coordinate field $\mathbf{Z}$. The CP-factorized 5D photometric LUT Φ_P maps $\mathbf{Z}$ to the initial enhanced state P^0 through pentalinear interpolation, denoted by $\mathcal{P}_{\Phi_P}$. Starting from P^0, CSD-LUT performs T successive channel-pair spatial residual refinements. Let P^t denote the RGB

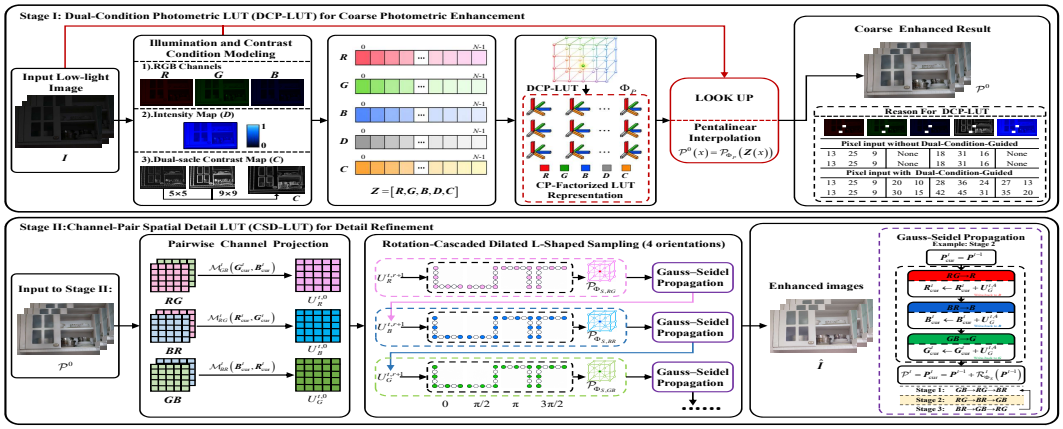


Figure 2: Overview of PD-Net: DCP-LUT performs photometric correction and CSD-LUT restores channel-correlated details.

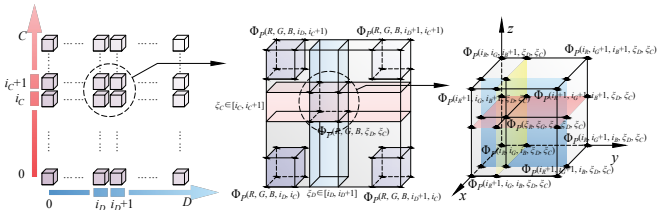


Figure 3: Illustration of Pentalinear Interpolation.

state after the t-th refinement and $\mathcal{R}^t_{\Phi_S}$ the corresponding residual lookup. The spatial LUT set Φ_S contains three CP-factorized pair-specific LUTs: $\Phi_{S,RG}$, $\Phi_{S,GB}$, and $\Phi_{S,BR}$. Each residual lookup combines rotation-cascaded dilated L-shaped lookup with Gauss-Seidel channel-pair updating. With $\mathcal{D}^T_{\Phi_S}$ denoting the complete T-stage spatial refinement, the overall PD-Net pipeline is

$$\hat{I} = \mathcal{D}^T_{\Phi_S}\left(\mathcal{P}_{\Phi_P}(R, G, B, D, C)\right). \tag{1}$$

2.2 Dual-Condition Photometric LUT (DCP-LUT)

DCP-LUT augments RGB indexing with spatial illumination and contrast cues, allowing pixels with similar RGB values but different photometric states to access distinct regions of Φ_P.

Illumination and Contrast Condition Modeling. A convolutional condition estimator $\mathcal{E}_\theta$, parameterized by θ, takes the input image I and predicts the illumination map D and dual-scale contrast map C. The illumination map D characterizes spatially varying exposure, whereas C adaptively fuses fine- and coarse-scale contextual features. Fine-scale context captures local variations around edges and textures, while coarse-scale context characterizes regional contrast and reduces sensitivity to isolated noise. At each pixel x, the lookup coordinate $\mathbf{Z}(x)$ comprises, in order, the values of $R(x)$, $G(x)$, $B(x)$, $D(x)$, and $C(x)$. The photometrically corrected state passed to CSD-LUT is

$$P^0 = \mathcal{P}_{\Phi_P}(\mathbf{Z}). \tag{2}$$

Pentalinear Interpolation. Because $\mathbf{Z}(x)$ is continuous and Φ_P is defined on a discrete five-dimensional lattice, we use pentalinear interpolation to retrieve its response. Each coordinate component is normalized to the unit interval. For the q-th component $Z_q(x)$, the continuous lattice coordinate $\xi_q(x)$, lower-node index $i_q(x)$, and fractional offset $\delta_q(x)$ are

$$\xi_q(x) = (N-1)Z_q(x), \quad i_q(x) = \min(\lfloor \xi_q(x) \rfloor, N-2), \quad \delta_q(x) = \xi_q(x) - i_q(x), \quad q = 1, \ldots, 5. \tag{3}$$

This construction keeps the lower-node index and its adjacent upper-node index within the valid range, even when $Z_q(x)$ reaches its upper bound of 1.

Let $\mathbf{i}(x)$ denote the five-dimensional vector formed by the lower-node indices $i_1(x), \ldots, i_5(x)$, which specifies the lower corner of the enclosing cell. Its 32 vertices are indexed by a binary displacement vector $\mathbf{c}$, whose five components $c_1, \ldots, c_5$ each take either 0 or 1. The interpolation weight assigned to the displaced vertex $\mathbf{i}(x) + \mathbf{c}$ is

$$\omega_{\mathbf{c}}(x) = \prod_{q=1}^{5} \left[(1 - c_q)\big(1 - \delta_q(x)\big) + c_q \delta_q(x)\right]. \tag{4}$$

The interpolated RGB response is

$$P^0(x) = \mathcal{P}_{\Phi_P}(\mathbf{Z}(x)) = \sum_{\mathbf{c}\in\{0,1\}^5} \omega_{\mathbf{c}}(x)\, \Phi_P[\mathbf{i}(x) + \mathbf{c}], \tag{5}$$

where the LUT entry $\Phi_P[\mathbf{i}(x)+\mathbf{c}]$ stores the RGB response at the corresponding vertex. Pentalinear interpolation is illustrated in Figure 3.

2.3 Channel-Pair Spatial Detail LUT (CSD-LUT)

Image structures show stronger cross-channel correlations than sensor noise (Yang et al., 2025). CSD-LUT exploits this distinction through complementary channel pairs, constraining spatial refinement with cross-channel information to suppress noise while preserving structural details. The three pair-specific 5D spatial LUTs in Φ_S are shared across all residual stages and orientations. After T stages, the ordered state $P^T = (R^T, G^T, B^T)$ is concatenated along channels by $\mathrm{Cat}_{\mathrm{ch}}$, yielding

$$\hat{I} = \mathcal{D}^T_{\Phi_S}(P^0) = \mathrm{Cat}_{\mathrm{ch}}\left(R^T, G^T, B^T\right). \tag{6}$$

Pairwise Channel Projection. Pairwise channel projection maps each complementary channel pair to a response, exposing cross-channel structural cues without expanding the 5D spatial index. At the t-th residual stage, the RG, GB, and BR pairs produce the target R, G, and B responses, respectively. Each pair is processed by a stage-specific learnable 2D LUT with bilinear interpolation, denoted by $\mathcal{M}^t_{RG}$, $\mathcal{M}^t_{GB}$, and $\mathcal{M}^t_{BR}$. Let $P^t_{\mathrm{cur}} = (R^t_{\mathrm{cur}}, G^t_{\mathrm{cur}}, B^t_{\mathrm{cur}})$ denote the current channel state within the t-th stage, initialized from P^{t-1}. The branchwise projections are defined as

$$U^{t,0}_R = \mathcal{M}^t_{RG}\left(R^t_{\mathrm{cur}}, G^t_{\mathrm{cur}}\right), \quad U^{t,0}_G = \mathcal{M}^t_{GB}\left(G^t_{\mathrm{cur}}, B^t_{\mathrm{cur}}\right), \quad U^{t,0}_B = \mathcal{M}^t_{BR}\left(B^t_{\mathrm{cur}}, R^t_{\mathrm{cur}}\right). \tag{7}$$

Here, $U^{t,0}_R$, $U^{t,0}_G$, and $U^{t,0}_B$ denote the projected responses of the RG, GB, and BR branches, respectively, before rotation-cascaded dilated L-shaped lookup. Five spatial samples from each projected response form the coordinate of the corresponding 5D spatial LUT in Φ_S.

2.3.1 Rotation-Cascaded Dilated L-Shaped Lookup.

After pairwise projection, a five-point dilated L-shaped pattern expands spatial coverage and recursively aggregates complementary structures across four orientations, with each response feeding the next. Let s denote the sampling interval, which is set to 2. In the reference orientation, the five offsets relative to the L-shaped corner are

$$\left\{\Delta^0_k\right\}_{k=1}^5 = \{(0,0),\ (s,0),\ (2s,0),\ (0,s),\ (0,2s)\}, \tag{8}$$

taking $\Delta^0_1 = (0,0)$ as anchor, the offsets at orientation $\theta_r = r\pi/2$ are defined as $\Delta^{\theta_r}_k = \mathrm{Rot}_{\theta_r}(\Delta^0_k)$, where $r = 0, 1, 2, 3$. Reflection padding is applied to samples outside the image domain.

For a channel-pair branch with target channel a and complementary channel b ($a, b \in \{R, G, B\}$ and $a \neq b$), the projected response $U^{t,0}_a$ is updated at orientation θ_r as

$$U^{t,r+1}_a(x) = \mathcal{P}_{\Phi_{S,ab}}\left(\left[U^{t,r}_a\left(x + \Delta^{\theta_r}_k\right)\right]_{k=1}^5\right). \tag{9}$$

Here, $\mathcal{P}_{\Phi_{S,ab}}$ denotes pentalinear interpolation over the pair-specific 5D spatial LUT $\Phi_{S,ab}$.

At each orientation, Eq. (9) is evaluated over all spatial locations, and the resulting response map $U_a^{t,r+1}$ becomes the input to the next orientation. After four recursive updates, $U_a^{t,4}$ serves as the spatial residual of target channel a.

Gauss-Seidel Propagation. To enable within-stage information propagation, the three channel pairs are updated sequentially, with each branch reading the latest state and immediately writing its residual to the target channel. At the beginning of the t-th residual stage, the current state is initialized as $P_{\text{cur}}^t = P^{t-1}$. For each ordered channel pair (a, b), pairwise channel projection followed by rotation-cascaded dilated L-shaped lookup produces the terminal response $U_a^{t,4}$ from P_{cur}^t. This response is added to target channel a, while the other two channels remain unchanged. The updated state is then used by the next channel pair, enabling structural information to propagate within the stage. To reduce order-induced channel bias, the processing order cycles through (GB, RG, BR), (RG, BR, GB), and (BR, GB, RG) across successive stages, distributing the three update positions evenly among the target channels.

After all three channel pairs have been updated, let $\mathcal{R}_{\Phi_S}^t$ denote the aggregate spatial residual produced by the sequential write-backs. The output of the t-th stage is

$$P^t = P_{\text{cur}}^t = P^{t-1} + \mathcal{R}_{\Phi_S}^t\left(P^{t-1}\right), \quad t = 1, \ldots, T. \tag{10}$$

2.4 CP-Factorized LUT Representation

A dense 5D LUT storing C_Φ responses per lattice node requires $C_\Phi N^5$ parameters. We therefore apply canonical polyadic (CP) decomposition (Fu et al., 2024) to express each response grid as a sum of rank-one tensors.

Let $\Phi \in \mathbb{R}^{N\times N\times N\times N\times N\times C_\Phi}$ denote a LUT. For response channel o, where $o = 1, \ldots, C_\Phi$, its scalar grid Φ_o is

$$\Phi_o = \sum_{k=1}^{K_\Phi} \mathbf{a}_{\Phi,k,o}^{(1)} \otimes \mathbf{a}_{\Phi,k,o}^{(2)} \otimes \mathbf{a}_{\Phi,k,o}^{(3)} \otimes \mathbf{a}_{\Phi,k,o}^{(4)} \otimes \mathbf{a}_{\Phi,k,o}^{(5)}, \tag{11}$$

where K_Φ is the CP rank, $\otimes$ denotes the vector outer product, and $\mathbf{a}_{\Phi,k,o}^{(q)} \in \mathbb{R}^N$ is the factor vector for the q-th indexing axis. The response at any lattice node is

$$\Phi[i_1, i_2, i_3, i_4, i_5, o] = \sum_{k=1}^{K_\Phi} \prod_{q=1}^{5} \mathbf{a}_{\Phi,k,o}^{(q)}[i_q]. \tag{12}$$

This factorization reduces the parameter count from $C_\Phi N^5$ to $5C_\Phi K_\Phi N$. For Φ_P, the count decreases from $3N^5$ to $15K_P N$; for the three scalar spatial LUTs Φ_S, it decreases from $3N^5$ to $15K_S N$. The CP factors are jointly optimized, and lattice-node responses are computed from factor products without materializing dense grids. Overall, the high-dimensional LUT parameter count decreases from $6N^5$ to $15K_P N + 15K_S N$.

2.5 Loss Function

PD-Net is trained end to end with an MSE loss $\mathcal{L}_{\text{mse}}$ for pixel reconstruction (Yang et al., 2025), an SSIM loss $\mathcal{L}_{\text{ssim}}$ for structural preservation (He et al., 2026), a low-frequency photometric consistency loss $\mathcal{L}_{\text{pho}}$ on P^0 (Lin et al., 2025), and LUT smoothness $\mathcal{L}_{\text{s}}$ and monotonicity regularization $\mathcal{L}_{\text{m}}$ (Fu et al., 2024). Thus,

$$\mathcal{L} = \mathcal{L}_{\text{mse}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{pho}}\mathcal{L}_{\text{pho}} + \lambda_{\text{s}}\mathcal{L}_{\text{s}} + \lambda_{\text{m}}\mathcal{L}_{\text{m}}, \tag{13}$$

where $\lambda_{\text{ssim}} = 0.1$, $\lambda_{\text{pho}} = 0.1$, $\lambda_{\text{s}} = 10^{-4}$, and $\lambda_{\text{m}} = 10$.

3 Experiments

3.1 Experimental Settings

Implementation Details. All experiments use PyTorch on an NVIDIA RTX 4060 Ti GPU. We train to convergence over 1×10^5 iterations using Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a batch size of 8, and an initial learning rate of 1×10^{-4}.

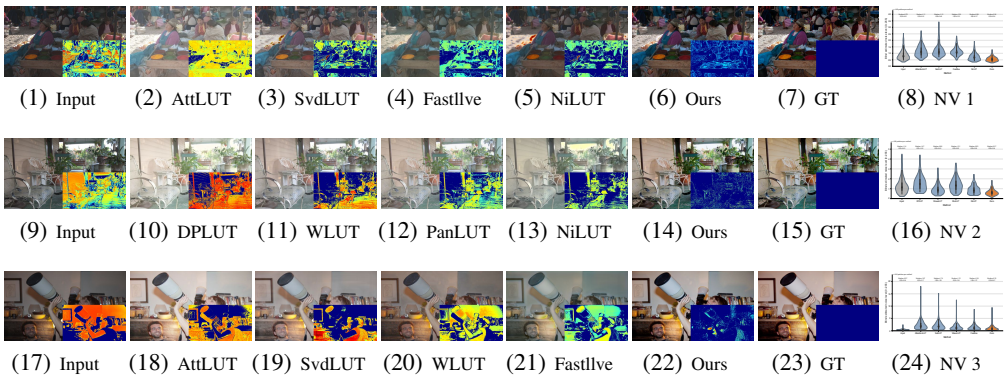


Figure 4: Visual comparison with recent SOTA methods on FiveK datasets. Best viewed zoomed in.

Datasets and Metrics. We evaluate PD-Net on paired and unpaired benchmarks. FiveK, LOLv1, LOLv2-real, and LOLv2-synthetic cover expert-retouched, real, and synthetic paired low-light images. DICM, MEF, LIME, NPE, and VV provide real unpaired indoor/outdoor scenes, while SICE provides paired multi-exposure images for ablation. PSNR, SSIM, and ΔE assess pixel accuracy, structure, and color on paired data; NIQE, BRISQUE (BRI), and PI assess no-reference perceptual quality Liu et al. (2026).

Comparison Methods. To evaluate PD-Net against recent state-of-the-art (SOTA) methods, we compare it with NiLUT (Conde et al., 2024), AttLUT (Fu et al., 2024), SvdLUT (Kim et al., 2025), Fastllve (Li et al., 2023), WLUT (He et al., 2026), DPLUT (Lin et al., 2025), FLUT (Hu et al., 2025), PanLUT (Cai et al., 2026), DnLUT (Yang et al., 2025), Darkir (Feijoo et al., 2025), DeepSPG (Lu et al., 2025), LatentD (Hu et al., 2026), TSCnet (Zhang et al., 2025), Adair (Cui et al., 2025), CIDNet (Zhao et al., 2025), SG-LLIE (Dong et al., 2025), LYT-NET (Brateanu et al., 2025), FFTF (Pei et al., 2025), CFWD (Xue et al., 2025), and WV-LUT (Li et al., 2025).

Table 1: Quantitative comparisons on FiveK test datasets. The best results are highlighted in bold.

Methods	#Para (K)	Runtime (ms)	Full Resolution(4k)			480p		
			PSNR↑	SSIM↑	ΔE↓	PSNR↑	SSIM↑	ΔE↓
NiLUT	133	8.3	26.35	0.925	7.32	26.87	0.944	7.28
AttLUT	593	6.8	25.56	0.926	7.53	25.61	0.932	7.48
SvdLUT	160.5	4.2	25.69	0.938	7.27	25.76	0.931	7.26
Fastllve	11000	13.5	26.22	0.861	7.44	26.54	0.878	7.39
WLUT	10742	10.7	25.75	0.879	7.47	25.96	0.885	7.43
DPLUT	**78**	6.9	20.56	0.738	7.97	21.33	0.822	7.92
PanLUT	662.6	3.8	26.24	0.911	7.58	26.52	0.938	7.36
Ours	848.9	**3.2**	**26.79**	**0.967**	**7.15**	**27.37**	**0.974**	**7.04**

3.2 Experimental Results

Quantitative Comparisons. On FiveK (Table 1), PD-Net achieved the best results in all six comparisons across the full-resolution and 480p settings. At 480p, it surpassed the next-best results by $0.50\,\mathrm{dB}$ in PSNR and 0.030 in SSIM, demonstrating stable reconstruction across resolutions. On the LOL benchmarks (Table 2), PD-Net ranked first in five of the six comparisons. Its PSNR exceeded the next-best results by $1.49\,\mathrm{dB}$, $1.10\,\mathrm{dB}$, and $0.87\,\mathrm{dB}$ on LOLv1, LOLv2-real, and LOLv2-synthetic, respectively. The only exception was SSIM on LOLv2-synthetic, which trailed CIDNet by just 0.002. Across the five no-reference benchmarks (Table 3), PD-Net ranked first in 11 of the 15 comparisons and second in the remaining 4. Specifically, its BRISQUE score on DICM was 0.720 lower than that of PanLUT. Its NIQE scores were also 0.373 lower than NiLUT on LIME and 0.195 lower than PanLUT on NPE. PD-Net achieved the best score for every metric on DICM, LIME, and NPE, demonstrating robust performance across datasets. These gains extend beyond brightness adjustment. DCP-LUT conditions the photometric mapping on illumination and contrast, allowing pixels with similar RGB values but different degradation states to receive distinct corrections.

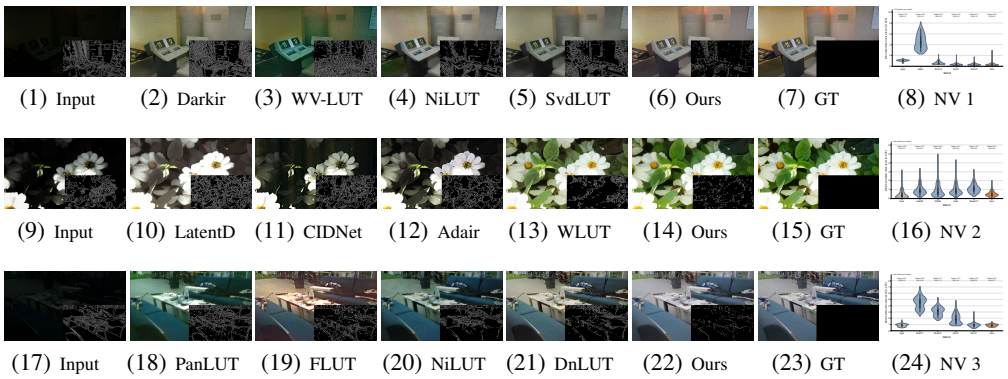


Figure 5: Visual comparison with recent SOTA methods on LOL datasets. Best viewed zoomed in.

CSD-LUT further uses cross-channel structural correlations and multidirectional neighborhoods to suppress residual noise and artifacts while preserving edges and textures. Together, the two stages improve photometric fidelity, structural recovery, and perceptual quality.

Table 2: Quantitative comparisons on LOL test datasets.

Methods	LOLv1		LOLv2-real		LOLv2-syn	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
NiLUT	23.23	0.794	25.62	0.836	25.44	0.928
WV-LUT	21.15	0.854	19.36	0.705	24.81	0.871
SvdLUT	24.25	0.878	24.68	0.844	21.41	0.848
Darkir	20.63	0.823	19.94	0.827	23.71	0.914
Adair	24.36	0.827	24.45	0.895	25.28	0.877
WLUT	25.32	0.922	25.49	0.889	24.68	0.919
CIDNet	23.81	0.857	24.11	0.868	25.13	**0.939**
LatentD	24.22	0.866	22.65	0.821	23.21	0.842
PanLUT	25.29	0.926	24.52	0.922	24.55	0.875
FLUT	23.73	0.854	24.26	0.863	25.15	0.920
DnLUT	24.88	0.886	25.55	0.875	25.68	0.925
Ours	**26.81**	**0.943**	**26.72**	**0.940**	**26.55**	0.937

Table 3: Quantitative comparisons on DICM,MEF,LIME,NPE and VV test datasets.

Methods	DICM			MEF			LIME			NPE			VV		
	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓	NIQE↓	BRI↓	PI↓
NiLUT	3.878	19.436	3.256	3.825	26.248	3.477	3.626	21.232	3.358	3.611	20.55	2.883	4.565	19.534	2.828
DPLUT	3.832	19.129	3.256	3.556	26.799	3.270	4.249	22.664	3.381	3.775	21.157	2.953	3.689	**14.153**	2.749
DnLUT	3.824	19.295	3.663	3.626	**20.277**	3.325	3.881	21.687	3.331	3.858	18.444	2.788	3.639	19.854	3.503
FFTF	4.519	27.922	3.700	3.608	26.716	3.286	4.463	25.170	3.376	4.124	28.887	3.534	5.312	22.800	3.648
PanLUT	3.545	18.878	3.344	3.658	24.404	3.367	4.133	20.451	3.441	3.546	22.235	3.425	**3.339**	22.24	3.374
CFWD	4.774	24.544	3.565	4.231	34.720	3.665	4.694	29.022	3.713	4.028	26.094	3.153	3.851	22.457	2.891
DeepSPG	4.422	20.662	3.826	4.745	28.577	4.233	4.575	27.425	4.874	4.365	26.542	3.768	3.835	24.411	2.770
TSCnet	4.277	22.369	3.538	3.969	22.775	**3.216**	4.110	21.528	3.562	4.211	22.563	3.525	3.824	22.663	3.467
LYT-NET	4.418	20.598	3.723	4.406	27.984	3.974	4.742	26.972	4.452	4.724	27.844	3.866	3.985	23.625	3.384
SG-LLIE	4.571	20.979	4.540	4.475	28.779	4.830	4.550	26.744	4.669	4.645	27.656	4.533	4.553	22.877	2.967
Ours	**3.421**	**18.158**	**3.152**	**3.436**	20.345	3.243	**3.253**	**20.154**	**3.155**	**3.351**	**18.322**	**2.686**	3.527	18.458	**2.719**

Qualitative Comparisons. Figures 4-6 compare color fidelity, structural preservation, intensity distributions, and residual noise. In Figure 4, the pseudo-color maps show pixel-wise errors relative to the GT. Blue denotes smaller errors, whereas yellow and red indicate larger luminance or color deviations. The SSIM difference maps in Figure 5 reveal local structural errors, fewer bright contours indicate better preservation of edges and textures. In Figure 6, the positions, ranges, and overlap of the RGB histogram curves characterize brightness, contrast, and color balance, respectively. Concentrations near 0 or 255 indicate intensity clipping. The rightmost noise violin (NV) and noise box (NB) plots summarize patch-wise blind noise estimates. Lower medians and narrower interquartile



Figure 6: Visual comparison with recent SOTA methods on DICM, MEF, LIME, and NPE.

ranges indicate lower residual noise and less variation across patches. On FiveK, PD-Net produces exposure, color, and local details that more closely match the GT, as shown by broader low-error regions and less noise distributions. On LOL, it reduces the color shifts, uneven exposure, and texture degradation observed in competing methods. The smaller structural errors and lower noise estimates further show that PD-Net preserves edges and textures while suppressing noise. On the unpaired benchmarks, PD-Net yields better-balanced RGB distributions, less accumulation at intensity boundaries, and lower, less noise estimates. These results reflect the complementary roles of the two stages. DCP-LUT reduces photometric errors through illumination- and contrast-conditioned mapping, whereas CSD-LUT uses cross-channel correlations and directional context to restore structural details and suppress channel-specific noise.

Table 4: Loss ablation on the SICE dataset.

Setting	$\mathcal{L}_{\text{mse}}$	$\mathcal{L}_{\text{ssim}}$	$\mathcal{L}_{\text{pho}}$	$\mathcal{L}_{\text{s}}$	$\mathcal{L}_{\text{m}}$	PSNR↑	SSIM↑	$\Delta E\downarrow$
MSE only	✓					26.42	0.951	7.67
W/O SSIM	✓		✓	✓	✓	26.95	0.959	7.32
W/O Photometric	✓	✓		✓	✓	26.72	0.964	7.49
W/O Smoothness	✓	✓	✓		✓	27.10	0.968	7.23
W/O Monotonicity	✓	✓	✓	✓		27.02	0.967	7.27
Full	✓	✓	✓	✓	✓	**27.21**	**0.969**	**7.18**

3.3 Ablation Study

Loss Functions. As shown in Table 4, the full loss achieved the best performance on the SICE dataset. The MSE loss provides pixel-level reconstruction supervision but does not explicitly constrain image structure, photometric consistency, or LUT regularity. Consequently, using the MSE loss alone decreased PSNR and SSIM by 0.79 dB and 0.018, respectively, and increased ΔE by 0.49. Removing the SSIM loss weakened the supervision of local luminance, contrast, and structural consistency, resulting in a 0.010 decrease in SSIM. The photometric consistency loss directly

Table 5: Architecture ablation study on the SICE dataset.

Setting	DCP-LUT			CSD-LUT				Metrics		
	Illumination D	Fine-scale contrast	Coarse-scale contrast	Pairwise projection	Rotation cascade	Gauss-Seidel propagation	Dilated L-shape	PSNR↑	SSIM↑	$\Delta E\downarrow$
Case 1		✓	✓	✓	✓	✓	✓	26.78	0.962	7.43
Case 2	✓			✓	✓	✓	✓	26.71	0.961	7.47
Case 3				✓	✓	✓	✓	26.39	0.955	7.65
Case 4	✓	✓		✓	✓	✓	✓	26.92	0.965	7.32
Case 5	✓		✓	✓	✓	✓	✓	26.86	0.964	7.36
Case 6	✓	✓	✓		✓	✓	✓	26.68	0.960	7.49
Case 7	✓	✓	✓	✓		✓	✓	26.83	0.963	7.38
Case 8	✓	✓	✓	✓	✓		✓	26.94	0.965	7.31
Case 9	✓	✓	✓	✓	✓	✓		27.02	0.967	7.26
Case 10	✓	✓	✓	✓	✓	✓	✓	**27.21**	**0.969**	**7.18**

constrains the low-frequency luminance and color distributions of the intermediate state P^0. Without this constraint, residual photometric errors propagate to the subsequent detail restoration stage, decreasing PSNR by 0.49 dB and increasing ΔE by 0.31. The smoothness regularization constrains variations between adjacent LUT entries and prevents abrupt local transitions. Removing it reduced PSNR by 0.11 dB and increased ΔE by 0.05. The monotonicity regularization preserves the relative intensity order between the input and output, reducing brightness inversions and color distortions. Its removal decreased PSNR by 0.19 dB and increased ΔE by 0.09. These results indicate that the loss terms provide complementary supervision for pixel reconstruction, structural preservation, photometric correction, and LUT mapping stability.

Network Architecture. To isolate each component, we evaluated nine controlled variants and the complete model on SICE. Cases 1-5 examine condition modeling in DCP-LUT. Case 1 removes illumination D, yielding a 4D LUT indexed by $[R, G, B, C]$. Case 2 removes contrast C, yielding a 4D LUT indexed by $[R, G, B, D]$. Case 3 removes both conditions, reducing DCP-LUT to an RGB 3D LUT. Cases 4 and 5 retain only the fine- or coarse-scale contrast branch to assess dual-scale complementarity. Cases 6-9 examine CSD-LUT components. Case 6 replaces pairwise channel projection with target-channel-only 1D projection to assess cross-channel information. Case 7 uses only the reference orientation to evaluate multi-directional spatial modeling. Case 8 replaces Gauss-Seidel propagation with synchronous channel updates, where all branches read the same stage input and write back jointly. Case 9 replaces dilated L-shaped lookup with its non-dilated counterpart to assess enlarged spatial coverage. Case 10 is the complete model.

Table 5 shows that illumination and contrast conditions provide complementary information. Removing D or C reduces PSNR by 0.43 dB and 0.50 dB, respectively. Removing both reduces PSNR and SSIM by 0.82 dB and 0.014, while increasing ΔE by 0.47. Retaining only the fine- or coarse-scale contrast branch also degrades performance, indicating that local texture variations and regional contrast are complementary. The CSD-LUT ablations show consistent gains from structural modeling. Replacing pairwise channel projection with target-channel-only projection reduces PSNR and SSIM by 0.53 dB and 0.009, and increases ΔE by 0.31, confirming the value of cross-channel correlations. Removing the rotation cascade, Gauss-Seidel propagation, or dilated sampling reduces PSNR by 0.38, 0.27, and 0.19 dB, reduces SSIM by 0.006, 0.004, and 0.002, and increases ΔE by 0.20, 0.13, and 0.08, respectively. These results verify the contributions of directional aggregation, within-stage channel propagation, and enlarged spatial coverage. The complete model performs best on all three metrics.

4 Conclusion

We presented PD-Net, a two-stage LUT framework that separates degradation-aware photometric correction from channel-correlated detail restoration. With context-predicted illumination and contrast, DCP-LUT guides RGB lookup for photometric correction. CSD-LUT further exploits complementary channel pairs, multi-directional dilated L-shaped neighborhoods, and Gauss-Seidel propagation to restore structures and suppress noise. Meanwhile, CP factorization reduces 5D mapping storage while preserving explicit interpolation-based inference. The serial dependencies remain a limitation. Future work will explore parallel propagation and video enhancement.

AI USE STATEMENT

Generative AI was not used in the conception of this study, method design, experimental analysis, or code implementation. Its use was limited to language assistance, including grammar correction, sentence refinement, and academic polishing. All AI-assisted revisions were reviewed and confirmed by the authors. The research ideas, technical methods, experimental results, analyses, and conclusions presented in this work were independently developed and verified by the authors. The authors take full responsibility for the scientific validity, accuracy, and final content of the manuscript, including any text refined with the assistance of generative AI.

REFERENCES

Alexandru Brateanu, Raul Balmez, Adrian Avram, Ciprian Orhei, and Cosmin Ancuti. Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement. *IEEE Signal Processing Letters*, 2025.

Zhongnan Cai, Yingying Wang, Hui Zheng, Panwang Pan, ZiXu Lin, Ge Meng, Chenxin Li, Chunming He, Jiaxin Xie, Yunlong Lin, et al. Pan-lut: Efficient pan-sharpening via learnable look-up tables. *Advances in Neural Information Processing Systems*, 38:126459–126483, 2026.

Marcos V Conde, Javier Vazquez-Corral, Michael S Brown, and Radu Timofte. Nilut: Conditional neural implicit 3d lookup tables for image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1371–1379, 2024.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In *International Conference on Learning Representations*, volume 2025, pp. 101306–101327, 2025.

Wei Dong, Yan Min, Han Zhou, and Jun Chen. Towards scale-aware low-light enhancement via structure-guided transformer design. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1460–1470, 2025.

Guodong Fan, Zishu Yao, Guang-Yong Chen, Jian-Nan Su, and Min Gan. Iniretinex: Rethinking retinex-type low-light image enhancer via initialization perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 2834–2842, 2025.

Daniel Feijoo, Juan C Benito, Alvaro Garcia, and Marcos V Conde. Darkir: Robust low-light image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 10879–10889, 2025.

Kang Fu, Yicong Peng, Zicheng Zhang, Qihang Xu, Xiaohong Liu, Jia Wang, and Guangtao Zhai. Attentionlut: Attention fusion-based canonical polyadic lut for real-time image enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3255–3259. IEEE, 2024.

Jinhong He, Minglong Xue, Wenhai Wang, and Mingliang Zhou. Optimizing a 4d lookup table for low-light video enhancement via wavelet priori. *IEEE Transactions on Multimedia*, 2026.

Liubing Hu, Chen Wu, Anrui Wang, Dianjie Lu, Guijuan Zhang, and Zhuoran Zheng. Flowlut: Efficient image enhancement via differentiable luts and iterative flow matching. *arXiv preprint arXiv:2509.23608*, 2025.

Renzhi Hu, Ting Luo, Gangyi Jiang, Leiming Liu, Yeyao Chen, Haiyong Xu, and Zhouyan He. Latentdark: Reflectance guided latent diffusion model for low-light image enhancement. *Signal Processing*, 238:110125, 2026.

Yan Huang, Xiaoshan Liao, Jinxiu Liang, Yuhui Quan, Boxin Shi, and Yong Xu. Zero-shot low-light image enhancement via latent diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3815–3823, 2025.

Wontae Kim, Keuntek Lee, and Nam Ik Cho. Lightweight and fast real-time image enhancement via decomposition of the spatial-aware lookup tables. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11895–11905, 2025.

Canlin Li, Haowen Su, Xin Tan, Xiangfei Zhang, and Lizhuang Ma. Wv-lut: Wide vision lookup tables for real-time low-light image enhancement. *IEEE Transactions on Multimedia*, 27:4441–4453, 2025.

Wenhao Li, Guangyang Wu, Wenyi Wang, Peiran Ren, and Xiaohong Liu. Fastllve: Real-time low-light video enhancement with intensity-aware look-up table. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8134–8144, 2023.

Yunlong Lin, Zhenqi Fu, Kairun Wen, Tian Ye, Sixiang Chen, Ge Meng, Yingying Wang, Chui Kong, Yue Huang, Xiaotong Tu, et al. Dplut: Unsupervised low-light image enhancement with lookup tables and diffusion priors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5316–5324, 2025.

Lin Liu, Junfeng An, Jianzhuang Liu, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, Yan Feng Wang, and Qi Tian. Low-light video enhancement with synthetic event guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 1692–1700, 2023.

Yuetong Liu, Yunqiu Xu, Yang Wei, Xiuli Bi, and Bin Xiao. Clear nights ahead: Towards multi-weather nighttime image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7395–7403, 2026.

Jialang Lu, Huayu Zhao, Huiyu Zhai, Xingxing Yang, and Shini Han. Deepspg: Exploring deep semantic prior guidance for low-light image enhancement with multimodal learning. In *Proceedings of the 2025 International Conference on Multimedia Retrieval*, pp. 935–943, 2025.

Yucong Meng, Zhiwei Yang, Kexue Fu, Zhijian Song, and Yonghong Shi. Image content matters: An image content aware state space model for accelerated mri reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 8025–8033, 2026.

Xiao Pei, Yongdong Huang, Weijian Su, Fengjuan Zhu, and Qiang Liu. Fftformer: A spatial-frequency noise aware cnn-transformer for low light image enhancement. *Knowledge-Based Systems*, 314:113055, 2025.

Zhen Wang, Yaozu Wu, Dongyuan Li, Shiyin Tan, and Zhishuai Yin. Thermal-aware low-light image enhancement: A real-world benchmark and a new light-weight model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8223–8231, 2025.

Minglong Xue, Jinhong He, Wenhai Wang, and Mingliang Zhou. Low-light image enhancement via clip-fourier guided wavelet diffusion. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(11):1–22, 2025.

Sidi Yang, Binxiao Huang, Yulun Zhang, Dahai Yu, Yujiu Yang, and Ngai Wong. Dnlut: Ultra-efficient color image denoising via channel-aware lookup tables. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7582–7591, 2025.

Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2020.

Miao Zhang, Jun Yin, Pengyu Zeng, Yiqing Shen, Shuai Lu, and Xueqian Wang. Tscnet: A text-driven semantic-level controllable framework for customized low-light image enhancement. *Neurocomputing*, 625:129509, 2025.

Gaoli Zhao, Kefei Zhang, Liangzhi Wang, Wenyi Zhao, and Weidong Zhang. Cidnet: Cross-scale interference mining detection network for underwater object detection. *Knowledge-Based Systems*, 324:113902, 2025.

BEYOND COLOR REMAPPING: PHOTOMETRIC–DETAIL NETWORK FOR LOW-LIGHT IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

1 DUAL-CONDITION PHOTOMETRIC LUT (DCP-LUT)

Details of Illumination and Contrast Condition Modeling. The illumination map D and the dual-scale contrast map C are predicted by a lightweight convolutional condition estimator $\mathcal{E}_\theta$. To provide pixel-wise condition maps while exploiting spatial context, $\mathcal{E}_\theta$ adopts a compact U-shaped encoder–decoder with skip connections. Given the normalized low-light image I, the encoder downsamples the feature maps and increases the channel width, producing a bottleneck feature at $1/32$ spatial resolution. The decoder then restores the resolution through transposed convolution and fuses encoder features through skip connections, yielding full-resolution feature Y_0 and coarse feature Y_2. The detailed architecture of $\mathcal{E}_\theta$, including the encoder, decoder, condition heads, and gated fusion, is summarized in Table 1.

The illumination head predicts D from Y_0 using a 1×1 convolution followed by a sigmoid function, so that D represents spatially varying exposure in the normalized range. The contrast head estimates fine- and coarse-scale contrast responses from different decoder levels. Specifically, the fine-scale response C_f is predicted from Y_0 using a 5×5 convolution to capture local variations around edges and textures, whereas the coarse-scale response C_c is predicted from Y_2 using a 9×9 convolution and upsampled to the input resolution to describe regional contrast. A spatially adaptive gate A then fuses the two responses:

$$A = \sigma(W_a * [C_f, C_c]) \,, \tag{1}$$

$$C = \sigma(W_o * (A \odot C_f + (1 - A) \odot C_c)) \,, \tag{2}$$

where W_a and W_o denote the learnable 1×1 convolution kernels for gate prediction and output projection, respectively. $*$ denotes convolution, $[\,\cdot, \cdot\,]$ channel-wise concatenation, $\odot$ element-wise multiplication, and $\sigma(\cdot)$ the sigmoid function. This gated fusion allows each pixel to adaptively select local or regional contrast cues. The resulting D and C are normalized to $[0, 1]$ and have the same spatial resolution as I. They are stacked with RGB values to form the 5D photometric coordinate $[R(x), G(x), B(x), D(x), C(x)]$. All parameters in $\mathcal{E}_\theta$ are optimized end to end with the LUT parameters.

Details of Pentalinear Interpolation. The pentalinear interpolation algorithm is presented in Algorithm 1. DCP-LUT performs photometric correction by interpolating the 5D coordinate $\mathbf{Z}(x)$. As defined in the article, this pixel-wise coordinate is

$$\mathbf{Z}(x) = [R(x), G(x), B(x), D(x), C(x)], \tag{3}$$

where the five dimensions correspond to the RGB values, illumination condition, and contrast condition. Since $\mathbf{Z}(x)$ is continuous and Φ_P is defined on a discrete five-dimensional lattice, the response is retrieved from the enclosing 5D cell. For the q-th component $Z_q(x)$, the continuous lattice coordinate $\xi_q(x)$, lower-node index $i_q(x)$, and fractional offset $\delta_q(x)$ are defined as

$$\begin{aligned} \xi_q(x) &= (N-1) Z_q(x), \\ i_q(x) &= \min(\lfloor \xi_q(x) \rfloor \,, N-2) \,, \\ \delta_q(x) &= \xi_q(x) - i_q(x), \qquad q = 1, \ldots, 5, \end{aligned} \tag{4}$$

this definition ensures that both $i_q(x)$ and $i_q(x)+1$ are valid lattice nodes, including the boundary case $Z_q(x) = 1$. Let $\bar{\delta}_q(x) = 1 - \delta_q(x)$. Let $\mathbf{i}(x) = [i_1(x), \ldots, i_5(x)]$ denote the lower corner of the enclosing cell. Its $2^5 = 32$ vertices are indexed by the binary displacement vector $\mathbf{c} = [c_1, \ldots, c_5] \in \{0, 1\}^5$. Following the main paper, the interpolation weight of vertex $\mathbf{i}(x) + \mathbf{c}$ is

$$\omega_{\mathbf{c}}(x) = \prod_{q=1}^{5} \left[(1 - c_q)\big(1 - \delta_q(x)\big) + c_q \delta_q(x) \right] . \tag{5}$$

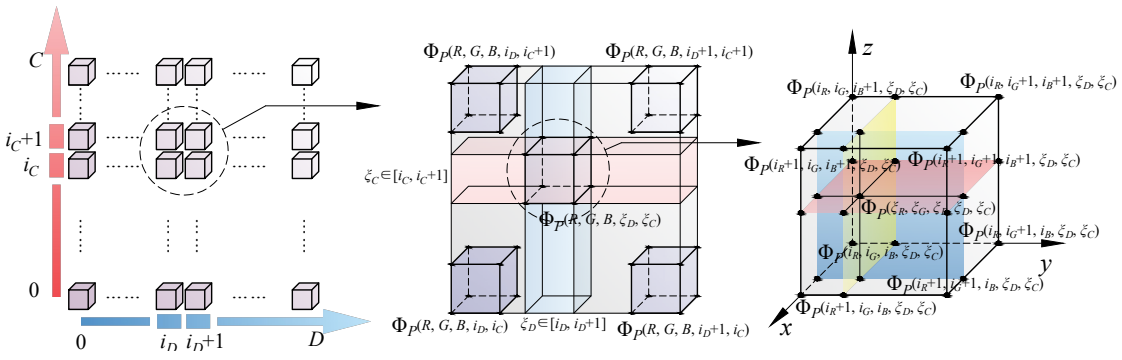


Figure 1: Illustration of Pentalinear Interpolation.

Table 1: Architecture of the condition estimator $\mathcal{E}_\theta$. Here, $n_f = 8$ is the base channel width, Y_2 and Y_0 denote the outputs of D3 and D5, respectively, and $\mathrm{Cat}_{\mathrm{ch}}$ denotes channel-wise concatenation. CB denotes Conv–LeakyReLU–IN, whereas DB denotes Deconv–LeakyReLU–IN.

Module	Input	Layer	Output size
Input	I	–	$3 \times H \times W$
E1	I	CB 3×3, $s = 2$	$n_f \times H/2 \times W/2$
E2	E1	CB 3×3, $s = 2$	$2n_f \times H/4 \times W/4$
E3	E2	CB 3×3, $s = 2$	$4n_f \times H/8 \times W/8$
E4	E3	CB 3×3, $s = 2$	$8n_f \times H/16 \times W/16$
E5	E4	CB 3×3, $s = 2$	$8n_f \times H/32 \times W/32$
D1	E5	DB 3×3, $s = 2$	$8n_f \times H/16 \times W/16$
D2	$\mathrm{Cat}_{\mathrm{ch}}(\mathrm{D1}, \mathrm{E4})$	DB 3×3, $s = 2$	$4n_f \times H/8 \times W/8$
D3 (Y_2)	$\mathrm{Cat}_{\mathrm{ch}}(\mathrm{D2}, \mathrm{E3})$	DB 3×3, $s = 2$	$2n_f \times H/4 \times W/4$
D4	$\mathrm{Cat}_{\mathrm{ch}}(\mathrm{D3}, \mathrm{E2})$	DB 3×3, $s = 2$	$n_f \times H/2 \times W/2$
D5 (Y_0)	$\mathrm{Cat}_{\mathrm{ch}}(\mathrm{D4}, \mathrm{E1})$	Deconv 3×3, $s = 2$ + LeakyReLU	$n_f \times H \times W$
D head	Y_0	Conv 1×1 + Sigmoid	$1 \times H \times W$
C_f branch	Y_0	Conv 5×5	$1 \times H \times W$
C_c branch	Y_2	Conv 9×9 + Up $\times 4$	$1 \times H \times W$
C fusion	$\mathrm{Cat}_{\mathrm{ch}}(C_f, C_c)$	Gate A + Conv 1×1 + Sigmoid	$1 \times H \times W$

The interpolated RGB response is therefore

$$P^0(x) = \mathcal{P}_{\Phi_P}(\mathbf{Z}(x)) = \sum_{\mathbf{c} \in \{0,1\}^5} \omega_{\mathbf{c}}(x)\, \Phi_P[\mathbf{i}(x) + \mathbf{c}]\,, \tag{6}$$

where $\Phi_P[\mathbf{i}(x) + \mathbf{c}]$ stores the RGB response at the corresponding lattice vertex.

Figure 1 illustrates the pentalinear interpolation process in DCP-LUT. The five axes of the photometric LUT are R, G, B, D, and C, and the continuous lookup coordinate is written as $(\xi_R, \xi_G, \xi_B, \xi_D, \xi_C)$. As shown in the figure, the interpolation is decomposed into two steps: condition-axis interpolation over D and C, followed by RGB-axis interpolation over R, G, and B.

First, the condition coordinate (ξ_D, ξ_C) lies inside the cell bounded by i_D, $i_D + 1$, i_C, and $i_C + 1$. It selects four neighboring RGB sub-cubes from Φ_P, corresponding to the four condition vertices (i_D, i_C), $(i_D + 1, i_C)$, $(i_D, i_C + 1)$, and $(i_D + 1, i_C + 1)$. Bilinear interpolation along the D and C axes combines these sub-cubes into the condition-fused RGB cube $\Phi_P(R, G, B, \xi_D, \xi_C)$, as marked in Figure 1. The condition weight is

$$\beta_{u,v}(x) = \left[(1-u)\bar{\delta}_D(x) + u\delta_D(x)\right] \times \left[(1-v)\bar{\delta}_C(x) + v\delta_C(x)\right], \quad u, v \in \{0, 1\}. \tag{7}$$

For any RGB lattice node $\mathbf{m} = [m_R, m_G, m_B]$, the condition-fused cube is

$$\Phi_P\left[\mathbf{m}, \xi_D, \xi_C\right] = \sum_{u,v \in \{0,1\}} \beta_{u,v}(x) \times \Phi_P\left[m_R, m_G, m_B, i_D + u, i_C + v\right]. \tag{8}$$

Second, the RGB coordinate (ξ_R, ξ_G, ξ_B) is interpolated inside $\Phi_P(R, G, B, \xi_D, \xi_C)$. This step produces the final response $\Phi_P(\xi_R, \xi_G, \xi_B, \xi_D, \xi_C)$, which corresponds to $P^0(x)$. The RGB interpolation weight is

$$\gamma_{p,q,r}(x) = \left[(1-p)\bar{\delta}_R(x) + p\delta_R(x)\right] \times \left[(1-q)\bar{\delta}_G(x) + q\delta_G(x)\right] \times \left[(1-r)\bar{\delta}_B(x) + r\delta_B(x)\right], \quad p, q, r \in \{0, 1\}. \tag{9}$$

Algorithm 1 Pentalinear Interpolation in DCP-LUT

Require: Coordinate field $\mathbf{Z}$, photometric LUT Φ_P, lattice resolution N
Ensure: Photometrically corrected state P^0

for each pixel x **do**
  $\mathbf{Z}(x) \leftarrow [R(x), G(x), B(x), D(x), C(x)]$
  for $q = 1, \ldots, 5$ **do**
    $\xi_q(x) \leftarrow (N-1)Z_q(x)$
    $i_q(x) \leftarrow \min(\lfloor \xi_q(x) \rfloor, N-2)$
    $\delta_q(x) \leftarrow \xi_q(x) - i_q(x)$
  end for
  $\mathbf{i}(x) \leftarrow [i_1(x), \ldots, i_5(x)]$
  $P^0(x) \leftarrow (0,0,0)$
  for each $\mathbf{c} \in \{0,1\}^5$ **do**
    $\omega_{\mathbf{c}}(x) \leftarrow 1$
    for $q = 1, \ldots, 5$ **do**
      $\omega_{\mathbf{c}}(x) \leftarrow \omega_{\mathbf{c}}(x) \left[(1-c_q)\big(1-\delta_q(x)\big) + c_q \delta_q(x)\right]$
    end for
    $P^0(x) \leftarrow P^0(x) + \omega_{\mathbf{c}}(x)\Phi_P[\mathbf{i}(x) + \mathbf{c}]$
  end for
end for
return P^0

The final response is therefore

$$P^0(x) = \Phi_P(\xi_R, \xi_G, \xi_B, \xi_D, \xi_C) = \sum_{p,q,r \in \{0,1\}} \gamma_{p,q,r}(x)\, \Phi_P\left[i_R + p, i_G + q, i_B + r, \xi_D, \xi_C\right]. \tag{10}$$

Substituting Eq. (8) into Eq. (10) gives the same 32-vertex pentalinear interpolation as Eq. (6). Thus, Figure 1 is a geometric decomposition of the same operation: DCP-LUT first interpolates the illumination–contrast-conditioned RGB sub-cubes and then interpolates the RGB coordinate inside the fused cube. This process allows pixels with similar RGB values but different illumination and contrast conditions to access different regions of Φ_P. Since Φ_P is CP-factorized, each queried lattice response is computed from factor products rather than from a materialized dense 5D grid. This changes only the storage and retrieval of LUT entries; the pentalinear interpolation rule remains unchanged.

1.1 Channel-Pair Spatial Detail LUT (CSD-LUT)

Details of Pairwise Channel Projection. Pairwise channel projection is used before spatial lookup to convert two correlated color channels into a single target-channel response. At the t-th residual stage, the current state is $P^t_{\text{cur}} = (R^t_{\text{cur}}, G^t_{\text{cur}}, B^t_{\text{cur}})$. The three ordered channel pairs RG, GB, and BR are assigned to the target channels R, G, and B, respectively. Thus, the RG branch predicts $U^{t,0}_R$, the GB branch predicts $U^{t,0}_G$, and the BR branch predicts $U^{t,0}_B$. Each branch uses a stage-specific learnable 2D LUT and bilinear interpolation. For the RG branch, the input coordinate at pixel x is $(R^t_{\text{cur}}(x), G^t_{\text{cur}}(x))$. Assume that each axis of $\mathcal{M}^t_{RG}$ contains $N_{\mathcal{M}}$ nodes. The continuous coordinates, lower-node indices, and fractional offsets are

$$\begin{aligned}
\eta_R(x) &= (N_{\mathcal{M}} - 1)R^t_{\text{cur}}(x), \\
j_R(x) &= \min(\lfloor \eta_R(x) \rfloor, N_{\mathcal{M}} - 2), \\
\alpha_R(x) &= \eta_R(x) - j_R(x), \\
\eta_G(x) &= (N_{\mathcal{M}} - 1)G^t_{\text{cur}}(x), \\
j_G(x) &= \min(\lfloor \eta_G(x) \rfloor, N_{\mathcal{M}} - 2), \\
\alpha_G(x) &= \eta_G(x) - j_G(x).
\end{aligned} \tag{11}$$

The bilinear weight of vertex $(j_R(x) + \mu, j_G(x) + \nu)$, where $\mu, \nu \in \{0,1\}$, is

$$\lambda^{RG}_{\mu,\nu}(x) = \left[(1-\mu)\big(1-\alpha_R(x)\big) + \mu\alpha_R(x)\right]\left[(1-\nu)\big(1-\alpha_G(x)\big) + \nu\alpha_G(x)\right]. \tag{12}$$

Algorithm 2 Gauss–Seidel Propagation in CSD-LUT

Require: Initial state P^0, residual stages T, spatial LUT set $\Phi_S = \{\Phi_{S,RG}, \Phi_{S,GB}, \Phi_{S,BR}\}$
Ensure: Refined RGB state P^T

for $t = 1, \ldots, T$ **do**
  $P^t_{\text{cur}} \leftarrow P^{t-1}$
  if $t \bmod 3 = 1$ **then**
    Use the processing order (GB, RG, BR)
  else if $t \bmod 3 = 2$ **then**
    Use the processing order (RG, BR, GB)
  else
    Use the processing order (BR, GB, RG)
  end if
  for each ordered channel pair (a, b) in the selected order **do**
    if $(a, b) = (R, G)$ **then**
      $U_R^{t,0} \leftarrow \mathcal{M}^t_{RG}(R^t_{\text{cur}}, G^t_{\text{cur}})$
      Obtain $U_R^{t,4}$ by rotation-cascaded dilated L-shaped lookup with $\Phi_{S,RG}$
      $R^t_{\text{cur}} \leftarrow R^t_{\text{cur}} + U_R^{t,4}$
    else if $(a, b) = (G, B)$ **then**
      $U_G^{t,0} \leftarrow \mathcal{M}^t_{GB}(G^t_{\text{cur}}, B^t_{\text{cur}})$
      Obtain $U_G^{t,4}$ by rotation-cascaded dilated L-shaped lookup with $\Phi_{S,GB}$
      $G^t_{\text{cur}} \leftarrow G^t_{\text{cur}} + U_G^{t,4}$
    else if $(a, b) = (B, R)$ **then**
      $U_B^{t,0} \leftarrow \mathcal{M}^t_{BR}(B^t_{\text{cur}}, R^t_{\text{cur}})$
      Obtain $U_B^{t,4}$ by rotation-cascaded dilated L-shaped lookup with $\Phi_{S,BR}$
      $B^t_{\text{cur}} \leftarrow B^t_{\text{cur}} + U_B^{t,4}$
    end if
  end for
  Let $\mathcal{R}^t_{\Phi_S}(P^{t-1})$ denote the aggregate residual produced by the sequential write-backs
  $P^t \leftarrow P^t_{\text{cur}} = P^{t-1} + \mathcal{R}^t_{\Phi_S}(P^{t-1})$
end for
return P^T

Using these weights, the projected response $U_R^{t,0}(x)$ of the RG branch is obtained by bilinearly combining the four neighboring LUT entries:

$$U_R^{t,0}(x) = \sum_{\mu,\nu \in \{0,1\}} \lambda^{RG}_{\mu,\nu}(x)\, \mathcal{M}^t_{RG}\left[j_R(x) + \mu,\, j_G(x) + \nu\right]. \tag{13}$$

The GB and BR branches are computed in the same way by replacing the input coordinates and projection LUTs with $(G^t_{\text{cur}}(x), B^t_{\text{cur}}(x))$, $\mathcal{M}^t_{GB}$, and $(B^t_{\text{cur}}(x), R^t_{\text{cur}}(x))$, $\mathcal{M}^t_{BR}$, respectively. This gives the three projected maps $U_R^{t,0}$, $U_G^{t,0}$, and $U_B^{t,0}$. The projection output remains single-channel for each branch. Therefore, the subsequent spatial LUT still receives five spatial samples from one projected response map as its 5D coordinate, rather than receiving multi-channel neighborhood samples. This preserves cross-channel structural cues through pairwise projection while keeping the dimensionality of the spatial LUT unchanged.

Analysis of Gauss-Seidel Propagation. Algorithm 2 makes the within-stage dependency explicit. Consider the order (GB, RG, BR). The GB branch first predicts $U_G^{t,4}$ from the initial state $P^t_{\text{cur}} = P^{t-1}$ and updates channel G. The subsequent RG branch therefore computes $U_R^{t,4}$ using the updated G, after which the BR branch computes $U_B^{t,4}$ using the updated R. The three residuals are thus conditionally coupled rather than independently predicted. Under synchronous propagation, all branches instead operate on the unchanged state P^{t-1}, so a correction produced by one branch cannot affect the others until the next residual stage. Gauss–Seidel propagation shortens this cross-channel propagation path by incorporating earlier updates into later residual predictions within the same stage. Cycling among (GB, RG, BR), (RG, BR, GB), and (BR, GB, RG) further assigns each target channel to every update position once per three-stage cycle, reducing dependence on a fixed branch order.

Algorithm 3 CP Vertex Reconstruction

Require: Lower corner $\mathbf{i}(x)$, binary displacement $\mathbf{c}$, factor vectors $\{\mathbf{a}^{(q)}_{\Phi,k,o}\}$, CP rank K_Φ, and response channels C_Φ
Ensure: Vertex response $\boldsymbol{\phi}_{\mathbf{c}}(x)$
for $o = 1, \ldots, C_\Phi$ **do**
  $\phi_{o,\mathbf{c}}(x) \leftarrow 0$
  for $k = 1, \ldots, K_\Phi$ **do**
    $\rho \leftarrow 1$
    for $q = 1, \ldots, 5$ **do**
      $\rho \leftarrow \rho\, \mathbf{a}^{(q)}_{\Phi,k,o}[i_q(x) + c_q]$
    end for
    $\phi_{o,\mathbf{c}}(x) \leftarrow \phi_{o,\mathbf{c}}(x) + \rho$
  end for
end for
return $\boldsymbol{\phi}_{\mathbf{c}}(x)$

2 CP-FACTORIZED LUTS.

The dense 5D LUT is not explicitly constructed during lookup. The cp vertex reconstruction procedure is summarized in Algorithm 3. For each vertex $\mathbf{i}(x) + \mathbf{c}$ required by pentalinear interpolation, its response is reconstructed directly from the corresponding CP factors. Let $\boldsymbol{\phi}_{\mathbf{c}}(x) = [\phi_{1,\mathbf{c}}(x), \ldots, \phi_{C_\Phi,\mathbf{c}}(x)]$ denote the response at this vertex. Its o-th component is

$$\phi_{o,\mathbf{c}}(x) = \sum_{k=1}^{K_\Phi} \prod_{q=1}^{5} \mathbf{a}^{(q)}_{\Phi,k,o}\,[i_q(x) + c_q]\,, \qquad o = 1, \ldots, C_\Phi. \tag{14}$$

The reconstructed response $\boldsymbol{\phi}_{\mathbf{c}}(x)$ replaces the corresponding dense entry $\Phi[\mathbf{i}(x) + \mathbf{c}]$ in the pentalinear interpolation.

For the photometric LUT Φ_P, $C_\Phi = 3$ and $K_\Phi = K_P$, so each reconstructed vertex stores an RGB response. For each pair-specific spatial LUT $\Phi_{S,ab} \in \Phi_S$, $C_\Phi = 1$ and $K_\Phi = K_S$, yielding a scalar response. The spatial CP factors are specific to the channel pair ab but shared across residual stages and rotation orientations.

Each vertex response is reconstructed, weighted by its pentalinear coefficient, and immediately accumulated into the output. It can then be discarded before the next vertex is evaluated. Thus, neither the dense 5D grid nor all 32 queried responses need to be stored. A lookup requires $O(2^5 \cdot 5K_\Phi C_\Phi)$ arithmetic operations and scales linearly with the CP rank and response channels. Because reconstruction consists only of factor products and summation, gradients propagate directly to the CP factors, which are jointly optimized with the remaining model parameters.

3 LOSS FUNCTION DETAILS

Let I^{gt} denote the ground-truth image and Ω the set of image pixels. The reconstruction loss directly supervises the final output $\hat{I}$ using the mean squared error (Yang et al., 2025):

$$\mathcal{L}_{\mathrm{mse}} = \frac{1}{3|\Omega|} \sum_{x \in \Omega} \left\| \hat{I}(x) - I^{\mathrm{gt}}(x) \right\|_2^2. \tag{15}$$

To preserve local luminance, contrast, and structural information, the SSIM loss (He et al., 2026) is defined as

$$\mathcal{L}_{\mathrm{ssim}} = 1 - \frac{1}{|\Omega|} \sum_{x \in \Omega} \mathrm{SSIM}_x \left(\hat{I}, I^{\mathrm{gt}} \right), \tag{16}$$

where SSIM_x evaluates structural similarity within a local window centered at x.

Because DCP-LUT performs photometric correction, we constrain the low-frequency content of its intermediate output P^0 (Lin et al., 2025). Let $\mathcal{B}$ denote the low-pass operator used during training and Ω_{LF} its output domain. The photometric consistency loss is

$$\mathcal{L}_{\mathrm{pho}} = \frac{1}{|\Omega_{\mathrm{LF}}|} \sum_{u \in \Omega_{\mathrm{LF}}} \left\| \mathcal{B}(P^0)(u) - \mathcal{B}(I^{\mathrm{gt}})(u) \right\|_1. \tag{17}$$

By suppressing high-frequency textures and noise, this loss emphasizes regional exposure and color consistency before spatial detail restoration. We also regularize the learnable LUTs to obtain stable mappings (Fu et al., 2024). Let $\lessgtr = \{\Phi_P, \Phi_{S,RG}, \Phi_{S,GB}, \Phi_{S,BR}\}$ denote the regularized 5D LUTs, $\mathbf{e}_q$ the unit displacement along the q-th indexing axis, and $\mathcal{N}_\Phi$ the set of valid neighboring lattice-node pairs. The smoothness loss is

$$\mathcal{L}_{\mathrm{s}} = \sum_{\Phi \in \lessgtr} \frac{1}{|\mathcal{N}_\Phi|} \sum_{(\mathbf{i},q)\in\mathcal{N}_\Phi} \|\Phi[\mathbf{i}+\mathbf{e}_q] - \Phi[\mathbf{i}]\|_2^2 . \tag{18}$$

It penalizes abrupt changes between adjacent LUT entries and promotes locally smooth interpolation. The monotonicity loss penalizes decreasing responses along each indexing axis:

$$\mathcal{L}_{\mathrm{m}} = \sum_{\Phi \in \lessgtr} \frac{1}{|\mathcal{N}_\Phi|} \sum_{(\mathbf{i},q)\in\mathcal{N}_\Phi} \|\mathrm{ReLU}\left(\Phi[\mathbf{i}] - \Phi[\mathbf{i}+\mathbf{e}_q]\right)\|_1 . \tag{19}$$

This constraint preserves the relative intensity order and reduces brightness inversions and color distortions. Because the 5D LUTs are CP-factorized, the neighboring responses in Eqs. (18) and (19) are computed directly from the CP factors without materializing dense grids. Overall,the complete training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{mse}} + \lambda_{\mathrm{ssim}}\mathcal{L}_{\mathrm{ssim}} + \lambda_{\mathrm{pho}}\mathcal{L}_{\mathrm{pho}} + \lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{s}} + \lambda_{\mathrm{m}}\mathcal{L}_{\mathrm{m}}, \tag{20}$$

where $\lambda_{\mathrm{ssim}} = 0.1$, $\lambda_{\mathrm{pho}} = 0.1$, $\lambda_{\mathrm{s}} = 10^{-4}$, and $\lambda_{\mathrm{m}} = 10$.

4 Experiments on the SMID and SDSD datasets

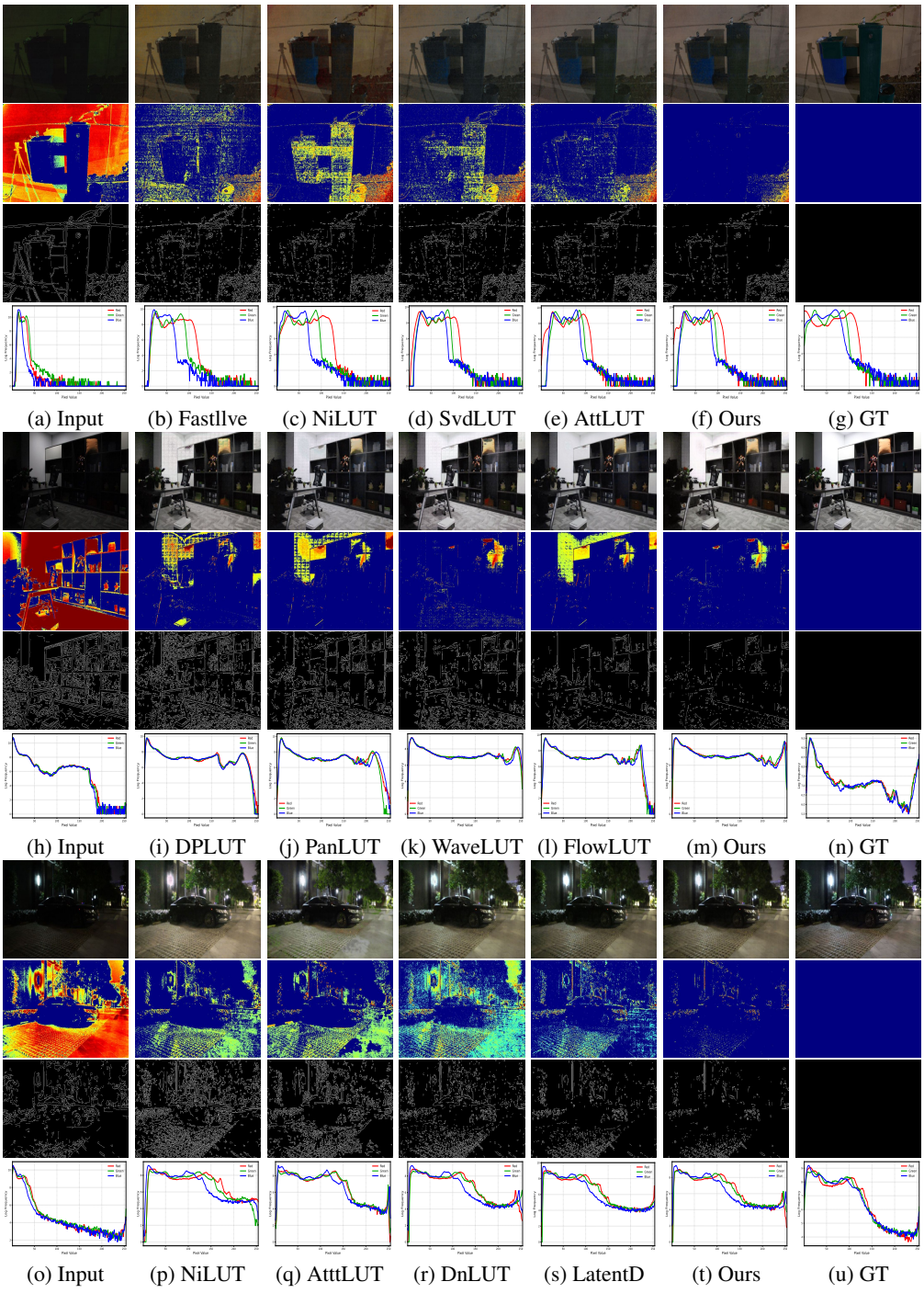
Quantitative Comparisons. As shown in Table 2, PD-Net ranks first in all six dataset–metric comparisons. Compared with the next-best result in each column, its PSNR improves by 0.20, 0.10, and 0.28 dB on SMID, SDSD-in, and SDSD-out, respectively. The corresponding SSIM gains are 0.016, 0.027, and 0.029. These results show that PD-Net improves both pixel-level accuracy and structural preservation rather than favoring a single metric. Most competing methods perform worse on SDSD-in than on SMID, indicating greater sensitivity to complex indoor lighting, local exposure variations, and scene structures. In contrast, PD-Net remains superior on both SDSD subsets and achieves its largest margins on SDSD-out. This result demonstrates its stronger adaptability to spatially varying outdoor illumination and complex structures. The larger SSIM gains on both SDSD subsets further indicate that its advantage extends beyond brightness correction to structural restoration.

The consistent PSNR gains mainly arise from the degradation-aware indexing of DCP-LUT. Methods based on RGB or intensity coordinates cannot distinguish pixels with similar colors but different illumination and contrast conditions. They may therefore produce under-enhancement, over-enhancement, or color deviations. DCP-LUT addresses this ambiguity using the 5D coordinate $\mathbf{Z} = [R, G, B, D, C]$, where D describes spatial exposure and C combines fine- and coarse-scale contrast. This coordinate distinguishes degradation states that RGB values alone cannot resolve, improving the accuracy and stability of photometric correction.

The SSIM gains on the SDSD subsets reflect CSD-LUT's structural modeling. Pairwise channel projection incorporates complementary channel information without increasing the dimensionality of the spatial LUT. It preserves cross-channel structures while reducing the influence of channel-specific noise. Rotation-cascaded dilated L-shaped lookup further aggregates context over four orientations and an enlarged spatial range. This design improves the recovery of edges, elongated contours, and irregular textures that pointwise or fixed-neighborhood lookup cannot adequately represent. Cyclic Gauss–Seidel propagation allows each branch to use the latest channel estimates generated within the same residual stage. Sequential write-back promotes cross-channel information transfer, while cyclic ordering reduces the channel bias caused by a fixed update sequence. These operations jointly improve photometric reconstruction and structural consistency. Overall, the performance gains arise from the complementary effects of degradation-aware photometric correction and channel-correlated spatial refinement.

Qualitative Comparisons. Figure 2 presents, from top to bottom, results on SMID, SDSD-in, and SDSD-out. Within each group, the first row shows the enhanced images, followed by pseudo-color error maps relative to the ground truth (GT), SSIM difference maps, and RGB histograms. In the pseudo-color maps, dark blue denotes small luminance and color errors, whereas green, yellow, and red indicate progressively larger deviations. Bright contours in the SSIM difference maps represent local structural errors; darker maps therefore indicate better preservation of edges and textures. The RGB histograms characterize intensity range, contrast, and channel balance. Greater agreement with the GT in peak position, distribution width, and channel overlap indicates more accurate photometric recovery, whereas concentrations near 0 or 255 suggest intensity clipping.

On SMID, Fastllve remains underexposed, while NiLUT, SvdLUT, and AttLUT introduce photometric deviations around the wall, container, and shadow transitions after brightening. These deviations appear as continuous green-to-red regions in the pseudo-color maps and pronounced object contours in the SSIM difference

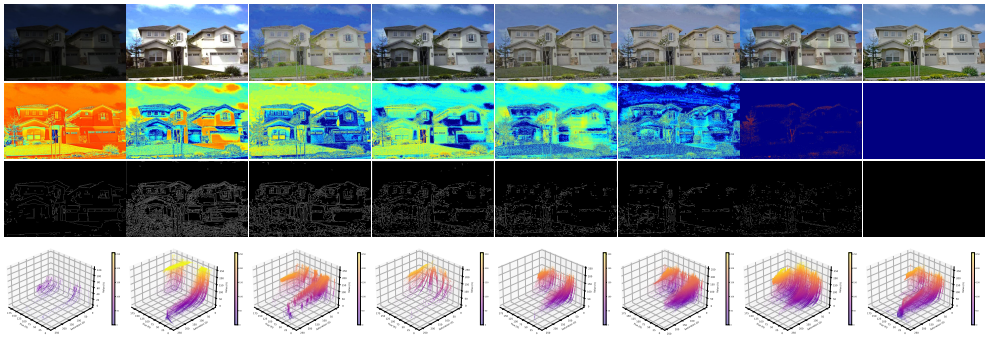


(a) Input (b) Fastllve (c) NiLUT (d) SvdLUT (e) AttLUT (f) Ours (g) GT

(h) Input (i) DPLUT (j) PanLUT (k) WaveLUT (l) FlowLUT (m) Ours (n) GT

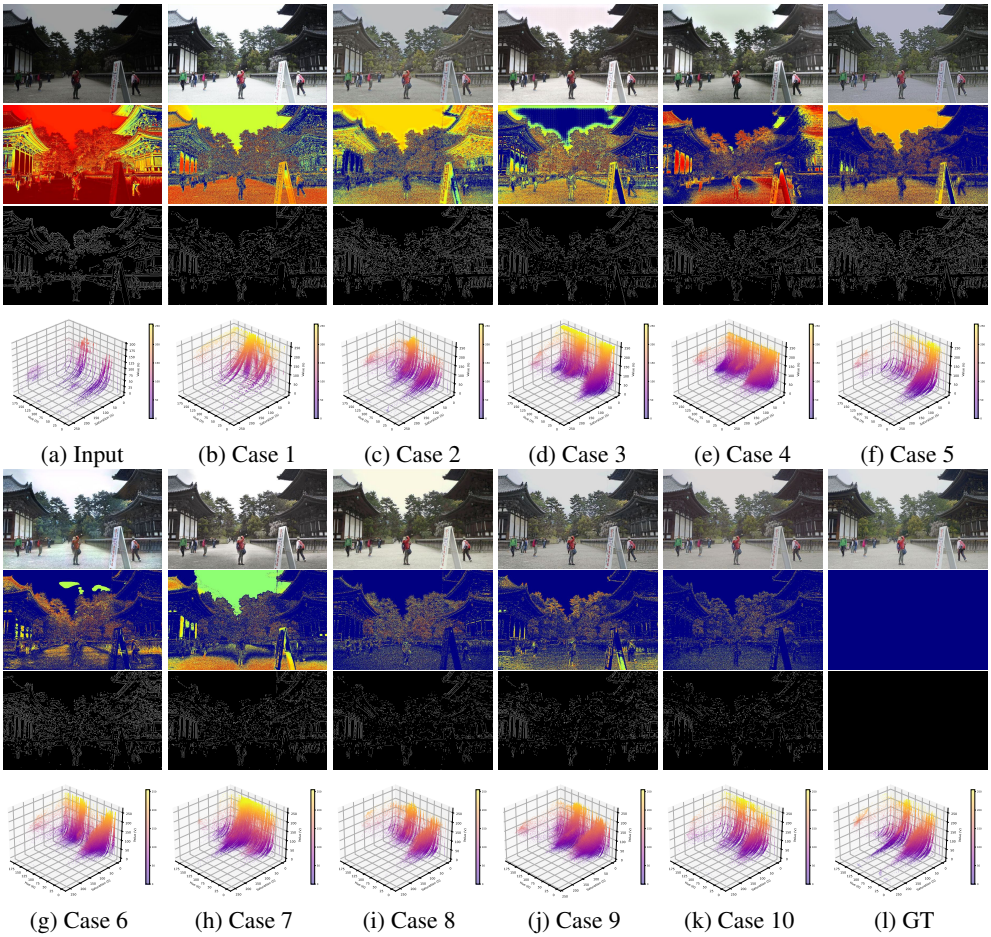
(o) Input (p) NiLUT (q) AtttLUT (r) DnLUT (s) LatentD (t) Ours (u) GT

Figure 2: Qualitative comparison results on the SMID,SDSD-in and SDSD-out datasets



(a) Input (b) $\mathcal{L}_{\text{mse}}$ Only (c) W/O $\mathcal{L}_{\text{ssim}}$ (d) W/O $\mathcal{L}_{\text{pho}}$ (e) W/O $\mathcal{L}_{\text{s}}$ (f) W/O $\mathcal{L}_{\text{m}}$ (g) Full (h) GT

Figure 3: Loss function ablation on the SICE dataset.



(a) Input (b) Case 1 (c) Case 2 (d) Case 3 (e) Case 4 (f) Case 5

(g) Case 6 (h) Case 7 (i) Case 8 (j) Case 9 (k) Case 10 (l) GT

Figure 4: Network Architecture ablation on the SICE dataset.

Table 2: Quantitative comparison on the SMID and SDSD test datasets. The best results are highlighted in bold.

Method	SMID		SDSD-in		SDSD-out	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
NiLUT Conde et al. (2024)	25.84	0.812	25.42	0.798	25.67	0.805
AttentionLUT Fu et al. (2024)	26.46	0.872	26.21	0.861	26.38	0.868
SvdLUT Kim et al. (2025)	25.93	0.841	25.68	0.829	25.86	0.836
Fastllve Li et al. (2023)	24.86	0.758	24.41	0.742	24.75	0.751
WaveLUT He et al. (2026)	27.03	0.908	26.94	0.901	26.98	0.905
DPLUT Lin et al. (2025)	25.41	0.793	25.16	0.781	25.33	0.788
FlowLUT Hu et al. (2025)	27.18	0.919	27.02	0.921	27.16	0.923
PanLUT Cai et al. (2026)	26.96	0.917	26.83	0.909	27.04	0.922
DnLUT Yang et al. (2025)	26.68	0.896	26.54	0.887	26.73	0.901
LatentD Hu et al. (2026)	27.14	0.937	27.11	0.916	27.09	0.927
Ours	**27.38**	**0.953**	**27.21**	**0.948**	**27.44**	**0.956**

maps. By contrast, PD-Net recovers wall tones, container colors, and shadow gradations that more closely match the GT. Its pseudo-color map is predominantly dark blue, with structural residuals confined to a few fine edges. Its RGB curves also better match the GT in peak position, distribution range, and decay pattern.

On SDSD-in, the errors of competing methods are concentrated around the shelves, tabletop, and local light sources. DPLUT, PanLUT, WaveLUT, and FlowLUT exhibit varying degrees of local exposure error or color imbalance, producing extensive green, yellow, or red regions in the pseudo-color maps and dense shelf contours in the SSIM difference maps. PD-Net improves overall visibility while retaining shadow gradations and preserving object and shelf boundaries. Its high-error regions are markedly smaller, its structural residuals are weaker, and its RGB curves more closely follow those of the GT. These results indicate that its brightness correction preserves local contrast and cross-channel relationships.

On SDSD-out, the pronounced exposure differences among the vehicle, road surface, and background place greater demands on spatially adaptive photometric correction. NiLUT, AttLUT, DnLUT, and LatentD retain broad photometric errors around the vehicle, road reflections, and dark background, with some results also exhibiting color shifts and weakened textures. The pseudo-color map of PD-Net remains predominantly dark blue, and its SSIM difference map contains the fewest bright contours, indicating better preservation of vehicle boundaries and road structures under complex outdoor illumination. Its RGB distribution also follows the GT more closely in both peak position and decay trend, without evident channel separation or boundary accumulation.

The three diagnostic maps further reveal the nature of these errors. Competing methods generally produce broad, spatially coherent pseudo-color residuals, indicating regional exposure or color-mapping errors rather than isolated pixel deviations. The residuals of PD-Net are sparser and largely confined to high-frequency edges, suggesting that most large-scale photometric errors have been removed. Its darker SSIM difference maps further indicate fewer contour displacements, halos, and over-smoothed textures. Meanwhile, the close agreement between its RGB histograms and those of the GT shows that the improvement does not arise from uniformly increasing brightness, but from jointly recovering the intensity range, local contrast, and channel balance.

These error patterns are consistent with the staged processing of PD-Net. DCP-LUT first corrects large-scale exposure, color, and contrast, producing an intermediate state P^0 with a stable low-frequency photometric distribution. CSD-LUT then targets the high-frequency errors associated with edges, textures, and channel-specific noise. This separation limits interference between low-frequency photometric adjustment and high-frequency structural restoration. Pentalinear interpolation further maintains continuous mappings across neighboring colors and degradation states, reducing banding and halos near local enhancement boundaries. Spatial refinement constrains residual prediction using cross-channel and neighborhood information, preserving consistent structural contours while suppressing isolated noise and unstable texture responses. The training objective further constrains these visual properties. The pixel reconstruction loss directly reduces point-wise deviations from the GT, consistent with the broader dark-blue regions in the pseudo-color maps. The SSIM loss preserves local luminance, contrast, and structure, corresponding to fewer bright contours in the difference maps. Applied to P^0, the low-frequency photometric consistency loss aligns regional exposure and color distributions before detail restoration, reducing broad coherent errors and improving agreement between the RGB histograms and the GT. LUT smoothness regularization discourages abrupt changes between adjacent entries, thereby reducing banding and boundary halos, while monotonicity regularization preserves relative intensity ordering and suppresses brightness reversals, channel inconsistencies, and intensity clipping. Together, these objectives align pixel accuracy, low-frequency photometry, local structure, and mapping stability, consistent with the smaller pixel errors, weaker structural residuals, and more accurate RGB distributions observed in Figure 2.

5 VISUALIZATION RESULTS OF ABLATION STUDY

Qualitative Analysis of Loss Functions. Figure 3 presents the visual results of the ablation study on the loss functions. The first row shows the enhanced images under different loss settings, followed by the pseudo-color error maps, SSIM structural error maps, and HSV color-space scatter plots. Each HSV plot maps image pixels into a three-dimensional space defined by hue H, saturation S, and value V. The position of the point cloud reflects overall color and brightness, its extent represents the color and contrast ranges, and its shape and density characterize their distributions. Greater agreement with the GT in position, orientation, and extent indicates more accurate recovery of hue, saturation, and brightness. Pronounced shifts, expansion, or compression instead suggest color deviations, oversaturation, exposure errors, or insufficient dynamic range. Because HSV scatter plots describe global color distributions without retaining pixel locations, they are interpreted jointly with the pseudo-color and SSIM error maps.

Using only $\mathcal{L}_{\mathrm{mse}}$ leaves broad luminance and color deviations across the building facades and sky. The pseudo-color map contains extensive high-error regions, while the SSIM map shows dense structural contours. The HSV cloud also differs from the GT in position and extent, indicating that pixel-wise reconstruction alone cannot constrain both global photometry and local structure. Without $\mathcal{L}_{\mathrm{ssim}}$, pronounced residuals remain around roofs, windows, and building boundaries, reflecting insufficient supervision of local luminance, contrast, and structure. Removing $\mathcal{L}_{\mathrm{pho}}$ produces spatially coherent photometric errors and shifts the value range and overall shape of the HSV cloud, showing that uncorrected low-frequency exposure and color errors in P^0 persist in the final output.

Without $\mathcal{L}_{\mathrm{s}}$ or $\mathcal{L}_{\mathrm{m}}$, errors concentrate around object boundaries, smooth gradients, and color transitions. Removing smoothness regularization produces uneven residuals near roofs, walls, and object edges, indicating discontinuous responses between adjacent LUT entries. Removing monotonicity regularization leaves persistent luminance and color errors along several boundaries. The outer extent of its HSV cloud also deviates from the GT, indicating less stable intensity ordering and color mapping.

With the complete objective, the pseudo-color map is predominantly dark blue except for a few narrow edges, the SSIM map contains the fewest bright contours, and the HSV cloud best matches the GT in position, shape, and extent. This show that $\mathcal{L}_{\mathrm{mse}}$, $\mathcal{L}_{\mathrm{ssim}}$, and $\mathcal{L}_{\mathrm{pho}}$ constrain pixel reconstruction, local structure, and low-frequency photometry, respectively, while $\mathcal{L}_{\mathrm{s}}$ and $\mathcal{L}_{\mathrm{m}}$ maintain mapping continuity and intensity ordering. Their joint optimization reduces regional exposure and color errors while preserving structural boundaries, local textures, and the color distribution.

Qualitative Analysis of Network Architecture. The architectural variants are defined as follows. Case 1 removes the illumination condition D and uses $[R, G, B, C]$ as the index. Case 2 removes the contrast condition C and uses $[R, G, B, D]$. Case 3 removes both conditions, reducing DCP-LUT to an RGB 3D LUT. Cases 4 and 5 retain only the fine- and coarse-scale contrast branches, respectively. Case 6 replaces pairwise channel projection with a target-channel-only 1D projection. Case 7 performs spatial lookup only in the reference orientation. Case 8 adopts synchronous channel updates, for which all branches read the same stage input and write their residuals back jointly. Case 9 replaces dilated L-shaped sampling with its non-dilated counterpart. Case 10 is the complete model.

Figure 4 presents the visual results of the ablation study on the network architecture. The DCP-LUT ablations affect photometric errors and HSV distributions. Without the illumination condition, Case 1 produces continuous yellow-to-red residuals over the tree canopy, building facades, and road. The center and extent of its HSV cloud along the V axis deviate from the GT, indicating that RGB and contrast alone cannot reliably adapt the mapping to exposure variations. Its SSIM map retains bright contours around roofs, branches, and road boundaries, showing that inaccurate exposure correction distorts structural contrast. Without the contrast condition, Case 2 produces broad yellow-green error bands along the sky-canopy boundary, building edges, and road. The altered shape and spread of its HSV cloud indicate that illumination alone cannot characterize local texture variations and regional contrast. Removing both conditions in Case 3 produces stronger spatially coherent errors and denser structural contours. This behavior reflects the ambiguity of RGB-only indexing, which assigns similar mappings to pixels with similar colors but different photometric states.

With only the fine-scale contrast branch, Case 4 reduces local residuals around roofs and foliage, but coherent low-frequency errors remain over the canopy, road, and shaded regions. Its HSV envelope remains misaligned with the GT, indicating that local cues alone cannot describe regional illumination changes. With only the coarse-scale branch, Case 5 suppresses broad exposure errors but retains pronounced structural residuals around foliage, roofs, and pedestrians. The distinct error patterns of these two variants show that fine-scale contrast describes local edges and textures, whereas coarse-scale contrast captures regional photometric variation. Their fusion is therefore required to correct both local and global contrast.

Cases 6–9 retain the complete DCP-LUT. Their HSV clouds consequently vary less than those of Cases 1–5, whereas their structural error maps differ more substantially, consistent with the role of CSD-LUT in spatial detail restoration. Without pairwise channel projection, Case 6 produces dense residuals around branches,

roofs, pedestrians, and road signs, indicating that a target channel alone cannot reliably distinguish cross-channel structures from channel-specific noise. Using only the reference orientation, Case 7 leaves directionally coherent residuals along the sky boundary, slanted roofs, and foliage, showing that a single sampling orientation cannot cover multi-directional structures. With synchronous updates, Case 8 confines most photometric errors to edge regions, but clear building and vegetation contours remain in the SSIM map. Because all branches operate on the same preceding state, corrections generated by one branch cannot inform subsequent branches within the current stage. With non-dilated sampling, Case 9 recovers short local edges but retains residuals over extended structures such as the canopy, rooflines, and road, indicating insufficient spatial coverage.

For Case 10, the pseudo-color errors contract to a few narrow boundaries, the SSIM map is nearly black, and the center, shape, and extent of the HSV cloud are most consistent with the GT. Together, the three visualizations show that illumination and dual-scale contrast conditions reduce regional exposure and color errors. Pairwise projection, rotation cascading, Gauss–Seidel propagation, and dilated sampling further provide cross-channel, multi-directional, within-stage, and large-range spatial context, respectively. The complete architecture therefore stabilizes the photometric distribution while minimizing local structural residuals.

AI USE STATEMENT

Generative AI was not used in the conception of this study, method design, experimental analysis, or code implementation. Its use was limited to language assistance, including grammar correction, sentence refinement, and academic polishing. All AI-assisted revisions were reviewed and confirmed by the authors. The research ideas, technical methods, experimental results, analyses, and conclusions presented in this work were independently developed and verified by the authors. The authors take full responsibility for the scientific validity, accuracy, and final content of the manuscript, including any text refined with the assistance of generative AI.

REFERENCES

Zhongnan Cai, Yingying Wang, Hui Zheng, Panwang Pan, ZiXu Lin, Ge Meng, Chenxin Li, Chunming He, Jiaxin Xie, Yunlong Lin, et al. Pan-lut: Efficient pan-sharpening via learnable look-up tables. *Advances in Neural Information Processing Systems*, 38:126459–126483, 2026.

Marcos V Conde, Javier Vazquez-Corral, Michael S Brown, and Radu Timofte. Nilut: Conditional neural implicit 3d lookup tables for image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1371–1379, 2024.

Kang Fu, Yicong Peng, Zicheng Zhang, Qihang Xu, Xiaohong Liu, Jia Wang, and Guangtao Zhai. Attentionlut: Attention fusion-based canonical polyadic lut for real-time image enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3255–3259. IEEE, 2024.

Jinhong He, Minglong Xue, Wenhai Wang, and Mingliang Zhou. Optimizing a 4d lookup table for low-light video enhancement via wavelet priori. *IEEE Transactions on Multimedia*, 2026.

Liubing Hu, Chen Wu, Anrui Wang, Dianjie Lu, Guijuan Zhang, and Zhuoran Zheng. Flowlut: Efficient image enhancement via differentiable luts and iterative flow matching. *arXiv preprint arXiv:2509.23608*, 2025.

Renzhi Hu, Ting Luo, Gangyi Jiang, Leiming Liu, Yeyao Chen, Haiyong Xu, and Zhouyan He. Latentdark: Reflectance guided latent diffusion model for low-light image enhancement. *Signal Processing*, 238:110125, 2026.

Wontae Kim, Keuntek Lee, and Nam Ik Cho. Lightweight and fast real-time image enhancement via decomposition of the spatial-aware lookup tables. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11895–11905, 2025.

Wenhao Li, Guangyang Wu, Wenyi Wang, Peiran Ren, and Xiaohong Liu. Fastllve: Real-time low-light video enhancement with intensity-aware look-up table. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8134–8144, 2023.

Yunlong Lin, Zhenqi Fu, Kairun Wen, Tian Ye, Sixiang Chen, Ge Meng, Yingying Wang, Chui Kong, Yue Huang, Xiaotong Tu, et al. Dplut: Unsupervised low-light image enhancement with lookup tables and diffusion priors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5316–5324, 2025.

Sidi Yang, Binxiao Huang, Yulun Zhang, Dahai Yu, Yujiu Yang, and Ngai Wong. Dnlut: Ultra-efficient color image denoising via channel-aware lookup tables. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7582–7591, 2025.